

Overview of the Open Provenance Model (OPM)

ESIP 2011 Winter Meeting – Preservation and Stewardship Cluster

Wednesday, January 5, 2011

Washington, DC

Hook Hua

Open Provenance Model Background



- International Provenance and Annotation Workshop (IPAW'06)
 - Session on provenance standardization and interoperability
- First Provenance Challenge (May 2006)
 - More informative than competitive
- Second Provenance Challenge
 - Establish interoperability of systems
- Third Provenance Challenge
 - Exchanging provenance encoded in OPM and answering queries
- Workshop in Salt Lake City (August 2007)
 - data model crafted and released as the Open Provenance Model (v1.00)

Open Provenance Model Requirements



1. Allow provenance exchange
 2. Tools for building and sharing common model
 3. Define provenance in a precise, technology-agnostic manner.
 4. Support representation of provenance for any “thing”
 5. Allow multiple levels of description to co-exist
 6. Define a core set of inference rules
- OPM was not intended to be fully like PML
 - OPM is more about *process provenance*
 - Identifiers: RDF/XML and OWL representations allows for URIs of resources

Layered Architecture

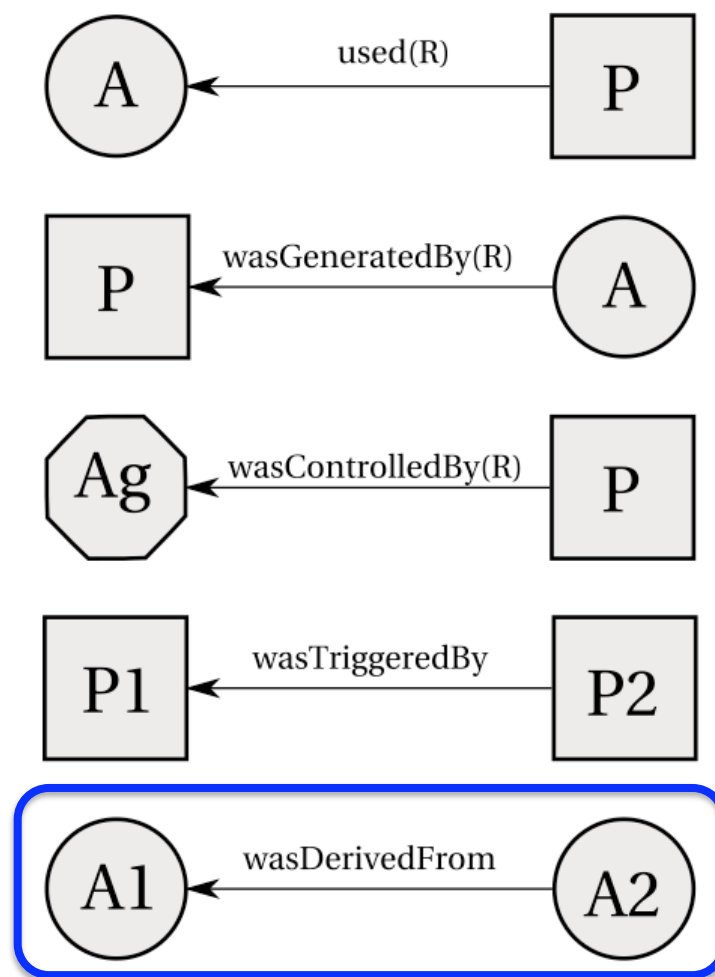
- Specifications have not been fully produced for all the layers



<http://openprovenance.org/>

- Nodes
 - Artifact
 - “Immutable piece of state, which may have a physical embodiment in a physical object, or a digital representation in a computer system.”
 - Process
 - “Action or series of actions performed on or caused by artifacts, and resulting in new artifacts.”
 - Agent
 - “Contextual entity acting as a catalyst of a process, enabling, facilitating, controlling, or affecting its execution.”
- A model of artifacts *in the past*, explaining how they *were* derived

- Edges (Dependencies)
 - OPM captures the causal dependencies between the artifacts, processes, and agents
 - Directed acyclic graph (DAG)
 - Annotations capturing additional information.
- Note direction of edge intended for past tense.
 - Invert for present tense
- Some dependencies may also be inferred with rules.



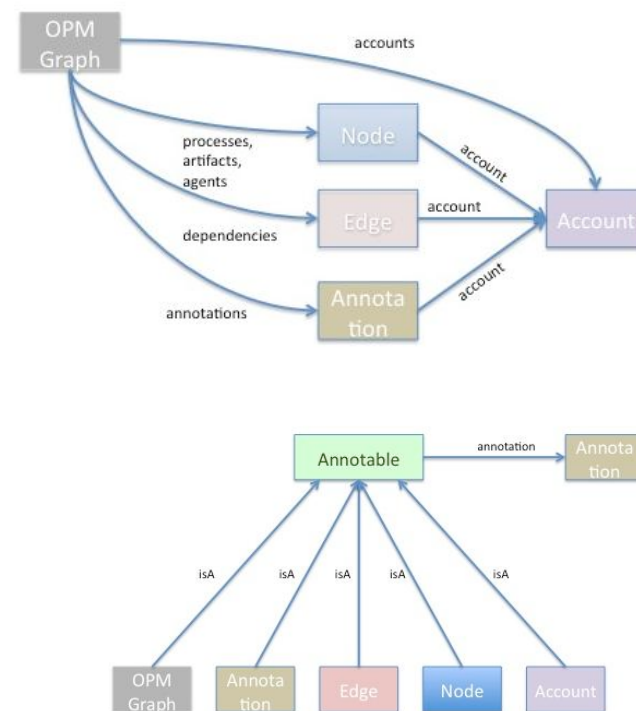
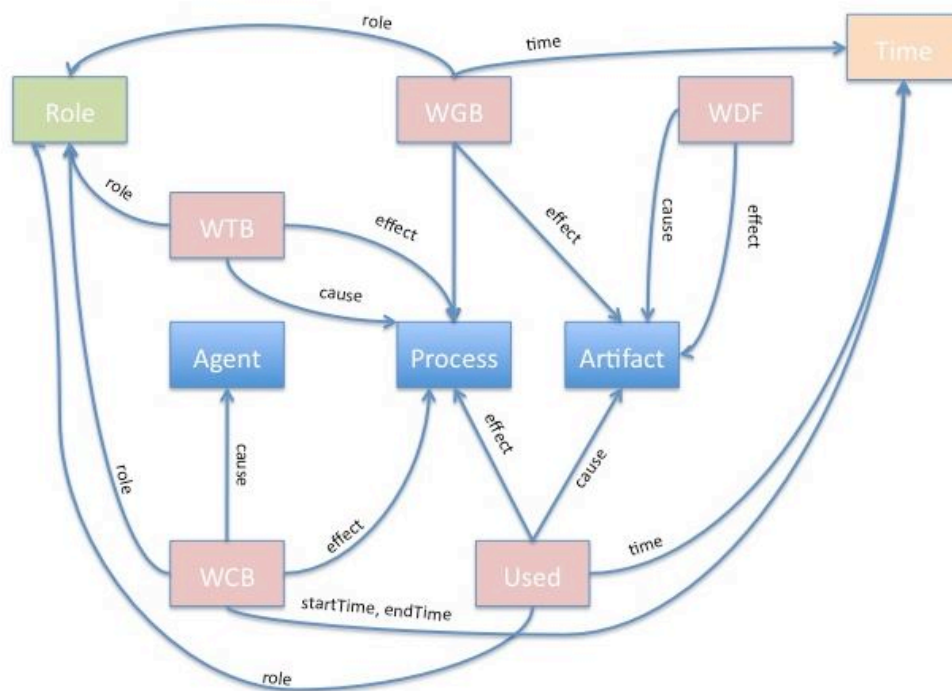
OPM Specifications



- **OPM:** The Open Provenance Model: Abstract Model
 - Open Provenance Model Core Specification (v1.1)
- **OPMX:** *The Open Provenance Model XML Schema*
- **OPMV:** *The Open Provenance Model Vocabulary*
- **OPMO:** *The Open Provenance Model OWL Ontology*
- **OPM4J:** *The Open Provenance Model Java Library*

OPMX: Open Provenance Model XML Schema

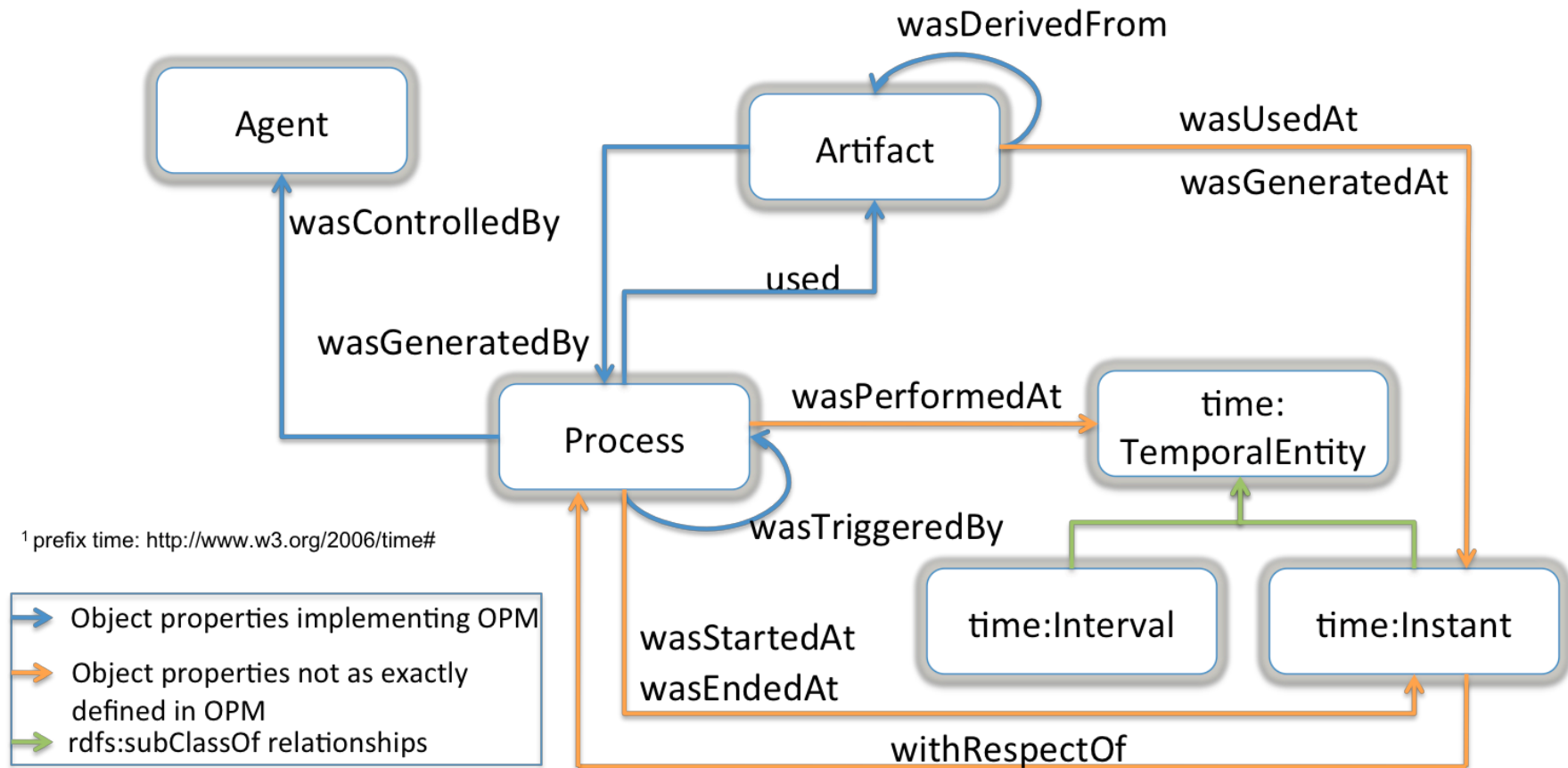
- Defines xsd types for abstract model entities
- Namespace
 - <http://openprovenance.org/model/opmx#>



<http://openprovenance.org/model/opmx>

- OPMV is a profile of OPM
- Can be used together with other provenance-related RDF/OWL vocabularies/ontologies
- Aims to reuse existing Semantic Web technologies and vocabularies as much as possible.
- Partitioned
 - Core ontology
 - Supplementary modules
- Time express as a *property* of Artifact and Process
 - Compared to OPM which expresses time as *annotations*

OPMV Core



<http://open-biomed.sourceforge.net/opmv/ns.html>

- Extends OPMV
- Defines classes and properties allowing all permitted inferences of the abstract model to be made.
 - Allows full expressivity and reasoning.
 - Allows inferences to be made
- RDF serialization of the OPM abstract model
- Time representation
 - *EventEdge (Used or WasGeneratedBy)*
 - An edge associated with a time instant
 - Statement: *EventEdge time OTme*
 - *WasControlledBy*
 - An Edge that can be associated with a time range.
 - Statements: *WasControlledBy {startTime,endTime} OTme*
 - *OTme (Observed Time)*
 - *exactlyAt, noEarlierThan, noLaterThan*
- Annotations
 - Statements
 - *Annotation property Property*
 - *ObjectProperty: Property key*
 - *DatatypeProperty: Property value*
- *Switch to Protégé visualization of OPM OWL* <http://openprovenance.org/model/opmo>

OPMV vs OPMO

Characteristic	OPMV	OPMO
Time Representation	As properties of Artifacts and Processes	As properties of Edges
Reasoning	OPM assertions	allows inferences.
OPM Edges	As properties	As classes. Allows all edge properties to be expressed, from which OPMV-style edges can be inferred automatically.
Named Graphs	Extensive use	Explicit class for accounts, and expresses account membership explicitly with a property
OPM Roles	Encodes implicitly in names of properties representing edges.	Roles explicit

References



- OPM Provenance Model (OPM)
 - <http://openprovenance.org/>
- OPMX: The Open Provenance Model XML Schema
 - <http://openprovenance.org/model/opmx>
- OPMV: The Open Provenance Model Vocabulary
 - <http://purl.org/net/opmv/ns>
- OPMO: The Open Provenance Model OWL Ontolog
 - <http://openprovenance.org/model/opmo>
- Luc Moreau, Ben Clifford, Juliana Freire, Joe Futrelle, Yolanda Gil, Paul Groth, Natalia Kwasnikowska, Simon Miles, Paolo Missier, Jim Myers, Beth Plale, Yogesh Simmhan, Eric Stephan, and Jan Van den Bussche. **The open provenance model core specification (v1.1)**. Future Generation Computer Systems, July 2010. ([doi: 10.1016/j.future.2010.07.005](https://doi.org/10.1016/j.future.2010.07.005)) (url: <http://eprints.ecs.soton.ac.uk/21449/>).